Deriving Design Recommendations Through Discount Usability Engineering: Ethnographic Observation and Thinking-Aloud Protocol in Usability Testing for Computer-Based Teaching Cases

Bonnie Kaplan, Ph.D., Yale Center for Medical Informatics, Yale University School of Medicine, New Haven, CT, USA

Usability engineering often involves experimental evaluation designs carried out in special usability laboratories. Though becoming more popular in medical informatics, the approach is little used. Possibly this is because of the expense involved in set-up, data collection, and data analysis. Excellent results may be obtained, however, by employing discount usability engineering and qualitative methods of data collection and analysis to generate recommendations to improve usability.

A prototype computer-based teaching case was evaluated in a discount usability engineering approach by combining modified ethnographic observation with simplified thinking aloud protocol. Data was collected and analyzed using standard approaches for qualitative data. This approach led to helpful recommendations for designing teaching cases. The project team believes this economical approach to usability testing may be helpful to others engaged in interface design.

INTRODUCTION

Approaches based on usability engineering for interface design recently have been introduced into medical informatics. Usability testing involves representative users involved in evaluating an interface and, thereby, contributing to an iterative process of interface design. There are different constituents of usability testing and different methods for conducting an assessment.[12,14] Usability testing also may include or be combined with a variety of methods to determine user satisfaction, ease of learning, times studies, and ease of use. More specific focus involves observation of individuals as they actually use the interface being tested. Frequently usability testing includes such hallmarks of cognitive analysis as thinking aloud protocols.[4, 9,14] Data also may be collected by different means, such as user surveys or videotapes of individuals as they use a system.

These methods have been used in a variety of medical informatics applications.[1,2,4,8,10,13,15] However, usability testing that follows a stringent approach is expensive, even if what a "modest"[14] usability laboratory is constructed and staffed. Elkin and colleagues,[2] for example, described the Usability Laboratory at the Mayo Clinic. It includes recording and monitoring equipment, cameras, microphones, computer equipment, control panels, sound-proof enclosures, and one-way mirrors. In addition to the laboratory, researchers competent in the methods of usability engineering are needed to conduct the studies. Perhaps it is not surprising, then, that Elkin and colleagues reported an "under-utilization" of the approach.

Discount usability engineering typically forgoes elaborate laboratory setups, experimental designs, videotaping, and detailed protocol analysis in favor of using user and task observation, scenarios, simplified thinking aloud, and heuristic evaluation.[12] Yao and Gorman[15] used discount usability engineering. They recommend it as more likely to be used than what usability experts generally propose.

Both Yao and Gorman and Elkin and colleagues advocate usability engineering through an experimental design to test interfaces. This paper reports a different approach in a formative initial usability evaluation of a prototype computer-based clinical case intended for educational purposes. The evaluation drew on discount usability engineering combined with modified ethnographic observation.[14] The approach quickly and economically generated valuable design insights without an elaborate set-up involving videotaping, multiple coders, and expensive statistical data analysis.

THE PROJECT

The Integrated Model of Aging and Geriatric Education (IMAGE), a Reynolds Project at the Yale School of Medicine, is an on-going effort to

integrate the process of geriatric care and the relevant knowledge of aging and geriatrics into clinical practice through education. A key element of the project is developing teaching cases to be used in computerized format for individual or group learning and as references for information needed in patient care. Future users of the teaching cases are expected to be medical students, residents, fellows, and other physicians seeking additional learning about geriatric care.

The first prototype teaching case included both information simulating multiple patient visits and also didactic resource material. The didactic material was displayed on the right-hand side of the screen and the medical case on the left. The case concerned an elderly man who was seen for three visits. For each visit, laboratory test results and medical record information were available by clicking on buttons or visuals of tabs. The clinical encounter was simulated in that learners clicked on buttons to select questions to interview the patient for a patient history and to examine the patient in a simulated physical. Answers to the questions were provided in text, images, videos, and numerics, as appropriate.

METHODS

Schneiderman[14] and Nielsen[12] each give guidelines for recruiting and informing participants and conducting usability tests. In medical informatics, Kushniruk, Patel, and Cimino[9] describe the following steps for user interface evaluations:

(1) development of test plan
(2) selection of representative users/study design
(3) selection of representative tasks/contexts
(4) setting up the test environment
(5) conducting the usability test (including introduction, conducting the test, and debriefing[11])
(6) data analysis
(7) recommendations to designers
(8) iterative input to design

The IMAGE evaluation was compatible with these general guidelines of usability testing, but differed in the means of data collection and analysis. Qualitative data collection and data analysis methods were used.

The approach drew on ethnographic observation, the goal of which Shneiderman describes as obtaining the necessary data to influence interface design. Ethnographic observation, like user and task observation, involves observing interfaces in use, and collecting both subjective and objective quantitative and qualitative data.[14] Instead of observing either in a specialized laboratory or in the actual work place, we observed individuals as they worked through the educational case on a desktop PC computer similar to one they eventually would use for educational cases. The observations were done in an ordinary room with a computer on a desk top. No other special set-up was used, and this environment was a close as possible to what the actual working environment would be for the system in use. One other difference, however, from the actual work environment, was that usability tests were scheduled and done with one individual at a time, whereas people using the system may use it wherever and whenever convenient, for as long as desired, and possibly with others present.

This modified ethnographic observation was combined with simplified thinking aloud protocol. Thinking aloud involves having the tester verbalize his or her thoughts while doing the assigned task. Nielsen describes the main difference between simplified and traditional thinking aloud: in simplified thinking aloud, analysis is based on the observer's notes instead of on videotapes. This approach is considered by software engineers as almost twice as useful as video protocols.[12]

Neilsen[11, 15] advises that 3-6 users are needed to identify prominent problems, and this number can obtain maximum cost to benefit ratios. An evaluator experienced in ethnographic methods observed eight first-year residents as each used the case. Five were medical residents and three were psychiatry residents. The primary purpose of the evaluation was to assess the educational aspects of the prototype case and to make design and usability recommendations in order to achieve educational goals.

Three different observation sessions were scheduled during which 2 or 3 residents in succession spent 45 minutes working through the case while thinking aloud. No training was provided. During the sessions, residents described what they were doing and explained it while doing it, sometimes volunteering information and sometimes in response to questions the evaluator asked. Those residents

who completed the case before the allotted 45 minutes had elapsed were interviewed (debriefed) briefly during the remaining time. Following the two observation sessions with the medical residents, the evaluator also was present at debriefing meetings their supervisor conducted with them to review the clinical aspects of the case.

Neither Nielsen nor Shneiderman discusses data collection or analysis methods. The evaluator used standard approaches for qualitative data. She took detailed notes of all these sessions. Notes included the navigational choices each resident made as he or she worked through the case, his or her comments while thinking aloud, both those volunteered and those given when questioned, the questions the evaluator asked, times when actions occurred, and remarks made during debriefing sessions. Wording was recorded as close to verbatim as possible.

Notes were analyzed according to four standard methods of qualitative data analysis:[5]

(1) Visual displays were created when the evaluator transcribed her notes, organizing them into tables showing each action taken, what time it was taken, what the medical resident said when he or she took those actions, and any computer problems that occurred. These tables were useful for following a medical resident's clinical thinking, reasons for navigational choices, reactions to what was presented, what information was sought that was not present, and how the program was working. Additional displays were compiled from these tables. One, for example, showed navigational decisions and choices residents would have liked. This display was used both for summary purposes and so that comparisons among residents could be made. Another display showed computer-related problems.

(2) Analytical memos--notes in which the evaluator records impressions, questions, or other reflective commentary--were made during and immediately after observations as well as during data analysis.

(3) Coding themes were identified through the constant comparative method.[3] Reports of the observation sessions were organized according to categories thus generated from the data.

(4) Narrative analysis was based both on the displays in which all comments were placed in context, as well as on the original notes.

Initial findings were detailed in a report and later summarized in presentations to the project team. The audience included case designers, residents' supervisors, other teaching faculty, and other project evaluators. The ensuing discussion provided validity checking.

FINDINGS

Residents seemed absorbed by the case and focused on it. Each resident immediately understood how to use the computer equipment and navigate through the case, though in some places they found the instructions confusing. They worked through the case quickly, and, for the most part, decisively. They had little trouble understanding the information that was displayed or what they were supposed to do. They liked it and thought it realistic. Residents got involved in it, wanted to know what happened in subsequent patient visits, wanted to complete the case if they had not been able to in the allotted time, wanted to know what they should have done, and discussed the case amongst themselves.

Findings concerning design for these kinds of teaching cases are summarized. These findings are grouped into those that might generalize to other displays, and those particular to this kind of teaching case.

General

On the whole, residents read down the part of the page pertaining to the visit. They read across each display line from left to right, much as if they were reading printed material. This had several consequences:

- Some made choices according to the order in which buttons were displayed.
- This may be why residents missed information. For example, some of them initially did not consult the patient's chart, and they did not notice that there was more information on subsequent screens.
- Likely because they were reading down the case, residents did not use didactic material displayed to the right of it, except at the ends

of visits. They did not use this material even when they did not know how to respond to aspects of the case. Three even looked up information elsewhere instead of consulting the materials included with the case.

- Residents wanted on-line feedback or an answer key. Their curiosity about the clinical aspects of the case was satisfied by debriefing sessions with their supervisor.
- Residents remarked that the screen design was "busy" and volunteered suggestions for improving it.

Particular

- Residents wanted more information about their role and the context of the visit. They sought clarification of whether they were seeing the patient as a primary care physician or as a referral, whether the patient had simply come in on his own or if this were an urgent visit, whether the patient were an in-patient or out-patient.
- Limitations on the number of questions the patient could be asked were annoying, though, as intended, they did force residents to make decisions. Most residents did not pay attention to the limits until they exceeded them, and then they counted their choices as they worked to make sure they did not exceed allotted limits again.
- The medical residents had difficulty with names of pharmaceuticals.
- Except for confusion as to whether they were functioning as specialists, there was little difference between Medicine and Psychiatry residents in how they executed the case. However, the number of residents was very small, making it difficult to be confident that these differences are not peculiar to the specific residents involved.

CONCLUSIONS

Forty-five minutes each with eight residents provided valuable information to designers of case content and screen layouts. All told, time spent on planning, executing, reporting, and presenting the evaluation was approximately 60 hours. The only costs were the evaluator's time, as no additional equipment or room space was needed for the observations, and the time of others on the project team involved in planning the evaluation and reviewing results.

By combining modified ethnographic observation with simplified thinking aloud protocol for data collection, and using standard methods of qualitative data analysis, the project team obtained a comprehensive analysis and set of recommendations far more useful than they had initially expected. These recommendations are serving as the basis for designing other teaching cases. Other cases will also undergo a similar usability evaluation. The project team, therefore, would not only advise this economical approach to usability testing, but also believes the recommendations this evaluation generated may be helpful to others engaged in designing computer-based teaching cases.

REFERENCES

1. Boyington A.R., Wildemuth B., Dougherty M.C. A usability engineering approach to development of a computer-based system for continence health promotion. Proc. AMIA Symp. 2002;:981.

2. Elkin P.L. Sorensen B. De Palo D. Poland G. Bailey K.R. Wood D.L. LaRusso N.F. Optimization of a research web environment for academic internal medicine faculty. Journal of the American Medical Informatics Association. 9(5):472-8, 2002 Sep-Oct..

3. Glaser B.G., Strauss A.L. The discovery of grounded theory: strategies for qualitative research. New York: Aldine De Gruyter, 1967.

4. Johnson C.M., Johnson T.R., Zhang J. Increasing productivity and reducing errors through usability analysis: a case study and recommendation. Proc. AMIA Symp. 2000;:394-398.

5. Kaplan B, Maxwell J.A. Qualitative research methods for evaluating computer information systems. In: Anderson J.G., Aydin C.E., Jay S.J., editors. Evaluating health care information systems; approaches and applications. Thousand Oaks, CA: Sage; 1994. p. 45-68.

6. Kinsey M., Cohn W., Knause W. The importance of heuristic evaluation and usability testing in the user interface design for a family health history web site. Proc. AMIA Symp. 2001;:945.

7. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Computers in Biology & Medicine. 32(3):141-9, 2002 May.

8. Kushniruk A.G., Kan M.-Y., McKeown K., Klavans J., Jordan D., LaFlamme M., Patel V.L. Usability evaluation of a text summarization system and three search engines: implications for the re-engineering of health care interfaces. Proc. AMIA Symp. 2002;:420-424.

9. Kushniruk A.W., Patel V.L., Cimino J.J. Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces. Proc. AMIA Symp. 1997;:218-222.

10. McDaniel A.M., Hutchinson S., Casper G.R., Ford R.T., Stratton R., Rembusch M. Usability testing and outcomes of an interactive computer program to promote smoking cessation in low income women. Proc. AMIA Symp. 2002;:-509-513.

11. Neilsen J. and Landaur T.K. A mathematical model of the finding of usability problems. Proc. ACM INTERCHI '93 Conf. 24-29 April, 1993, Amsterdam, The Netherlands, pp. 206-213.

12. Nielsen, J. Usability engineering. New York: AP Professional, Academic Press, 1993.

13. Rukab J.A., Walji M.F. Reducing errors and improving performance by applying usability engineering principles to a web-based prototype of an integrative medicine reference tool. Proc. AMIA Symp. 2002;:1146.

14. Shneiderman B. Designing the user interface: strategies for effective human-computer interaction. Third ed. Reading, Mass.: Addison-Wesley; 1998.

15. Yao P., Gorman P.N. Discount usability engineering applied to a interface for web-based medical knowledge resources. Proc. AMIA Symp. 2000;:928-932.

Acknowledgements:
The evaluation was funded by The Reynolds Foundation.

Contact information:
Bonnie Kaplan, Ph.D.
President, Kaplan Associates
59 Morris St
Hamden, CT 06517
USA
bonnie.kaplan@yale.edu